MODIS retrievals of cloud effective radius in marine stratocumulus exhibit no significant bias

Mikael K. Witte[1,2], Tianle Yuan[3,4], Patrick Y. Chuang[2], Steven Platnick[4], Kerry G. Meyer[4], Gala Wind[4,5], Haflidi H. Jonsson[6]

[1]National Center for Atmospheric Research*, Boulder, CO, USA
[2]Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, CA, USA
[3]Joint Center for Earth Systems Technology, University of Maryland at Baltimore County, MD, USA
[4]Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[5]Science Systems and Applications, Inc., Lanham, MD, USA
[6]Naval Postgraduate School, Monterey, CA, USA

Key Points:

- No systematic bias is found comparing MODIS cloud effective radius with in situ measurements taken by a phase Doppler interferometer.
- The lack of bias observed in precipitating conditions is consistent with modeling studies.
- Agreement between MODIS and in situ measurements depends on the ability of in situ probes to measure wide drop size distributions.

*The National Center for Atmospheric Research is supported by the National Science Foundation

Corresponding author: M. K. Witte, mkwitte@ucar.edu

Abstract

Satellite retrievals of cloud effective radius r_e are frequently validated using aircraft in situ measurements. Past intercomparisons have found a significant bias toward larger remotely sensed r_e. Explanations for this bias have focused on retrieval algorithms and large-scale heterogeneity, with in situ measurement uncertainty regarded as a minor factor. We compare Moderate Resolution Imaging Spectroradiometer (MODIS) r_e with in situ observations of marine stratocumulus clouds from three aircraft campaigns using a phase Doppler interferometer (PDI) probe. Retrieved and in situ r_e typically agree within uncertainty in both non-precipitating and drizzling conditions with no apparent systematic bias (mean bias of -0.22 μm, mean relative bias 3%). Agreement depends on the choice of in situ probe as well as microphysical context. We demonstrate that probes must adequately characterize the width of the drop size distribution to avoid systematic underestimation of r_e.

1 Introduction

Satellites offer the only global-scale characterization of cloud properties such as cloud effective radius r_e and cloud optical depth. From these basic cloud properties, quantities such as cloud drop number concentration and liquid water path may be derived. The Moderate Resolution Imaging Spectroradiometer (MODIS) cloud product (MOD/MYD06, *Platnick et al.* [2017]) is the current state of the art and has been widely used in the research community to study cloud properties, Earth's radiation budget, and aerosol-cloud interactions (e.g. *Quaas et al.* [2009]; *Su et al.* [2010]; *Yuan et al.* [2011]; *Andersen et al.* [2017]; *Bennartz and Rausch* [2017]). Assessing the accuracy of the satellite retrieval algorithms requires independent constraints. Aircraft in situ measurements of the cloud drop size distribution (DSD) are often used for such intercomparisons, from which r_e, cloud optical depth and other bulk quantities (cloud drop number concentration, liquid water path, rain rate, etc.) can be calculated. Since the earliest validation studies using aircraft observations [*Nakajima et al.*, 1991; *Platnick and Valero*, 1995], intercomparisons have shown that satellite and in situ r_e are highly correlated but with a persistent bias toward higher satellite-retrieved r_e. A number of more recent in situ validation studies using aircraft observations over the eastern Pacific Ocean [*Painemal and Zuidema*, 2011; *Zheng et al.*, 2011; *Min et al.*, 2012; *King et al.*, 2013; *Noble and Hudson*, 2015] continue to report that retrieved r_e is systematically greater than in situ measurements (mean bias 1.75-2 μm, *Painemal and Zuidema* [2011]; *Min et al.* [2012]; *Noble and Hudson* [2015]) and that the bias can primarily be attributed to retrieval uncertainty with a minimal contribution from uncertainty in the aircraft observations. In other words, the bias is caused by a systematic overestimate in remotely-sensed r_e as opposed to an underestimate by the in situ measurements. Despite that each of the aforementioned studies used a unique instrumentation configuration, very few (e.g. *Platnick and Valero* [1995] and *King et al.* [2013]) address the important point that different cloud probes may obtain differing results, whether due to differences in operating principles, measurement range or aircraft sampling strategy.

The underlying theory for satellite retrievals contains inherent assumptions about clouds and under many circumstances these assumptions are violated, which can lead to biases. The potential sources of error have been explored in a large body of literature and can be broadly divided into four categories: 1) spatial heterogeneity, including 3D radiative effects and pixel-scale horizontal variability [*Marshak et al.*, 2006; *Kato and Marshak*, 2009; *Zhang et al.*, 2012]; 2) assumptions about the shape of the DSD, including the role of vertical structure and precipitation [*Nakajima et al.*, 2010; *Min et al.*, 2012; *Kokhanovsky et al.*, 2013]; 3) view geometry dependence [*Maddux et al.*, 2010; *Grosvenor and Wood*, 2014; *Liang et al.*, 2015]; and 4) error in ancillary products such as retrieved cloud top pressure [*Painemal and Zuidema*, 2011; *Baum et al.*, 2012].

Improvements in satellite retrieval algorithms over time must also be taken into consideration. For this study, we use MODIS Collection 6 retrievals [*Platnick et al.*, 2017] while

previous east Pacific in situ validation studies have used Collection 5.1 or earlier. Of greatest interest to this work, Collection 6 has improved cloud top pressure and temperature retrievals and new bidirectional reflectance lookup tables for oceanic pixels that, on average, result in somewhat decreased r_e compared to Collection 5.1 (by ~1 μm at most, *Rausch et al.* [2017]) over the two east Pacific stratocumulus decks that will be examined here.

The goal of this study is to compare aircraft in situ microphysical observations collected by PDI with MODIS Collection 6 retrievals of r_e and revisit the results of recent intercomparison studies (in particular, *Zheng et al.* [2011] and *Noble and Hudson* [2015]) to re-evaluate the assertion of *Painemal and Zuidema* [2011] that instrumentation error can only account for up to 30% of the previously observed discrepancy between retrieved and in situ r_e.

2 Observations and methods

2.1 Aircraft observations

Measurements are taken from three different field campaigns focused on sampling marine stratocumulus, all flown by the Center for Interdisciplinary Remotely-Piloted Aircraft Studies (CIRPAS) Twin Otter aircraft: the Marine Stratus/Stratocumulus Experiment (MASE) [*Lu et al.*, 2007], the Physics of Stratocumulus Top experiment (POST) [*Carman et al.*, 2012; *Gerber et al.*, 2013], and the Variability of American Monsoon Systems (VAMOS) Ocean-Cloud-Atmosphere-Land Study (VOCALS) [*Zheng et al.*, 2011; *Mechoso et al.*, 2014]. Each campaign is briefly described below including a general overview of aircraft sampling strategy. Relevant aircraft instrumentation for each campaign is given in the next section.

MASE Twin Otter (TO) flights occurred during July 2005 in the northeastern Pacific near Monterey, California in a box roughly bounded by 36° to 37° N and 122° to 123° W. Note that the MASE measurements referenced in *Noble and Hudson* [2015] were taken aboard the US Department of Energy Gulfstream-1 [*Wang et al.*, 2009] which flew near Point Reyes, California, to the north of the TO sample area. Of the 13 research flights during MASE, we analyze four daily flights from 14-17 July 2005 when PDI was operational. Research flights sampled the cloud deck in a series of level legs from below cloud base to near cloud top, flown roughly perpendicular to the mean wind while the aircraft drifted with the mean flow. Near-cloud top legs were flown within ~20 m of cloud top and averages over these legs will be used for the intercomparison.

POST took place during July and August 2008 slightly farther offshore than the MASE sample area, in a box bounded by 35.5° to 37.5° N and 122.5° to 124.5° W. We analyze all 8 daytime flights during POST. The flight plans again were intended to drift with the mean flow and approximately follow a Lagrangian air mass, but sampling in the vertical was quite different than MASE. Instead of level legs, the purpose of POST was to extensively sample the interface between the stratocumulus layer and the free troposphere by repeatedly flying a sawtooth pattern from ~100 m below cloud top to ~100 m above.

VOCALS was a multi-platform campaign that took place over the southeastern Pacific during October and November of 2008 off the coast of Chile, of which the TO was one of five aircraft deployed. While numerous past satellite-in situ intercomparisons have utilized VOCALS observations [*Painemal and Zuidema*, 2011; *Min et al.*, 2012; *King et al.*, 2013; *Kokhanovsky et al.*, 2013], only *Zheng et al.* [2011] analyzed TO observations. Because of the limited range of the TO with respect to the other aircraft involved in VOCALS, flights were limited to the near-coastal region near 20° S, 72° W [*Zheng et al.*, 2011]. We exclude RF1-2, RF6, RF9 and RF15-18 because of instrument and data system issues. The standard flight pattern during VOCALS TO flights was similar to the level legs flown during MASE and near-cloud top legs are used for comparison with satellite retrievals.

2.2 In situ instrumentation

The PDI (Artium Technologies, Inc., Sunnyvale, CA, USA) effectively measures drops 2-100 μm in diameter. For a full description of the airborne probe and a direct comparison of in situ PDI observations with other cloud probes during MASE, see *Chuang et al.* [2008]. PDI deduces drop size by measuring the phase shift of the scattered interference pattern caused by a drop passing through the intersection of two laser beams. Beyond the TO, PDI has been deployed as the facility-standard drop size resolving probe for the helicopter-towed Airborne Cloud Turbulence Observation System (ACTOS, *Siebert et al.* [2006]; *Henrich et al.* [2010]; *Ditas et al.* [2012]; *Siebert et al.* [2013]; *Siebert and Shaw* [2017]), the ongoing NASA ORACLES mission and at a mountaintop station in the German Alps [*Siebert et al.*, 2015]. Of particular relevance to this study, *Henrich et al.* [2010] use ACTOS to compare in situ observations of continental stratiform clouds with MODIS retrievals, although their sampling method is problematic for comparison with MODIS because the clouds were quite thin (geometric thickness $<$ 100 m) and the vertical location of the helicopter tow package within cloud was highly uncertain. PDI has also been used in numerous laboratory experiments to study cloud droplet formation and evolution (e.g. *Ruehl et al.* [2008, 2010, 2012]; *Saw et al.* [2012]; *Chang et al.* [2016]; *Chandrakar et al.* [2016]) and industrial and engineering multiphase flows (combustion, spray measurement, etc.; [cf. *Bachalo*, 2000; *Tropea*, 2011; *Albrecht et al.*, 2013]).

Several other cloud probes are operated on the TO from which effective radius can be obtained: (i) Droplet Measurement Technologies (DMT; Longmont, CO) Cloud and Aerosol Spectroradiometer (CAS, measures drops 0.35-50 μm in diameter, *Baumgardner et al.* [2001]; no measurements available from MASE); (ii) Particle Measuring Systems (Boulder, CO) Forward Scattering Spectroradiometer Probe (FSSP, 1-46 μm, *Pinnick et al.* [1981]); (iii) Gerber Scientific, Inc. (Reston, VA) Particulate Volume Monitor (PVM, 5-50 μm, *Gerber et al.* [1994]); and (iv) DMT Cloud Imaging Probe (CIP) for drizzle-sized drops (25-1550 μm). DMT packages the CAS and CIP probes together as CAPS (Cloud, Aerosol and Precipitation Spectrometer); we use both probes to calculate r_e following *Noble and Hudson* [2015] and will refer to the CAS and CIP combination as CAPS. No CAPS measurements are available from MASE. We use liquid water content calculated from CIP DSDs (LWC_{CIP}) as a proxy for drizzle intensity. For comparison, instruments deployed aboard other aircraft used for MODIS-in situ intercomparisons is given in the supporting information [*Lance et al.*, 2010].

Beyond its minimal sizing error (~1 μm; [*Chuang et al.*, 2008; *Henrich et al.*, 2010]), the PDI has several other advantages over other commonly used in situ probes. PDI is not susceptible to mistaking the coincidence of small drops as a single larger drop, which would otherwise distort the size distribution, especially at larger sizes. PDI data can also be used to infer the view volume as a function of size, which is critical to accurate size distribution measurement. For example, *Chuang et al.* [2008] show that the view volume difference between 10 and 60 μm drops can be a factor of $\sim$ 2. Instead of determining drop size from the intensity of scattered light, PDI sizing instead depends on wavelength which leads to a significantly reduced need for calibration and a built-in mechanism for rejecting coincident or non-spherical particles, which generate different interference patterns than single spherical particles [*Chuang et al.*, 2008; *Albrecht et al.*, 2013]. Furthermore, sizing calibration can be independently verified by comparing PDI and aircraft measurements of true air speed [*Baumgardner et al.*, 2011]. Finally, PDI measures a broader size range compared to other cloud droplet probes flown on the TO: CAS, FSSP and PVM measure drops up to $d \sim 50$ μm in diameter while PDI nominally measures drops up to $d \sim 100$ μm. As such, only PDI measurements are used to calculate reference r_e. Using merged DSDs from PDI and CIP (using only drops $d > 62$ μm from CIP) has no significant effect on r_e (differences less than 0.5 ţm) in the near-cloud top region relevant to the comparison with MODIS retrievals because very few drizzle particles are encountered there.

Confidence in PDI measurements can be gained by comparing integrated PDI size distributions with LWC observations from the PVM. Note that the PVM uses a completely dif-

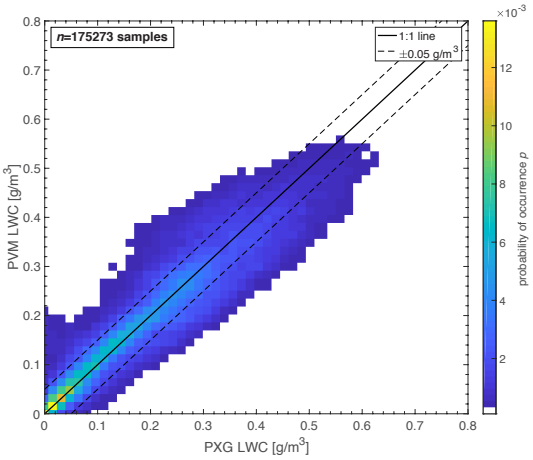

Figure 1. Joint probability density function of PVM-emulating PDI (PXG) versus PVM LWC. See text for details regarding calculation of PXG LWC. Number of samples n is the total number of observations for which PDI and PVM concurrently measure LWC $>$ 0.005 g/m^3. Shading is truncated at $p = 1 \times 10^{-4}$ for visual clarity.

ferent operating principle than the PDI; it measures populations of drops (rather than single drops) via near-forward scattering. To compare LWC from both instruments, we first must account for the response of the PVM to drops as a function of size [*Gerber*, 1991; *Wendisch et al.*, 2002; *Chuang et al.*, 2008]. Based on wind tunnel experiments, *Wendisch et al.* [2002] proposed shifting the manufacturer-provided PVM size response function by 20-30 µm in diameter. We apply the 30 µm shifted size response function to calculations of LWC from PDI DSDs and compare all measurements from MASE, POST and VOCALS in Fig. 1. The slope of the robust best fit line is 0.845 ± 0.001 with $r^2 = 0.87$ (compared to slope = 0.665 ± 0.001, $r^2 = 0.85$ for unmodified PDI LWC), which independently demonstrates that the third moment of the PDI size distributions is of very high fidelity. This provides confidence in the PDI-derived effective radii since this quantity depends on the second and third moments of the size distribution.

All measurements were processed to a common 1 Hz data stream, corresponding to a spatial resolution of approximately 55 m, and all cloud probes report size-resolved measurements except the PVM, which measures integral moments of the DSD. Manufacturer processing algorithms were used for the CAPS (CAS and CIP) probes and all probes were calibrated prior to, during and after each campaign. We subsample points with PDI liquid water content (LWC) greater than 0.01 g m^{-3} for the MODIS intercomparison to ensure that we analyze only cloudy air. Choosing a different LWC threshold does not alter the results. To calculate aircraft r_e for comparison with MODIS, we use full leg averages for level legs and sawtooth maneuvers since they are flown in close proximity to cloud top. Level and sawtooth legs were typically 10 minutes in duration (or $\sim$ 30 km in length). When sawtooth profiles are used, we average over the top 30 m in cloud.

2.3 Satellite retrievals and intercomparison method

We use MODIS level 2, Collection 6 [*Platnick et al.*, 2017] retrievals from both the Aqua and Terra platforms of r_e at 2.1 µm (Cloud_Effective_Radius product) and H_σ, a measure of subgrid variability (Cloud_Mask_SPI product, plane 2). This study does not consider r_e retrievals from other bands. We use a method similar to that of *Min et al.* [2012] to subsample MODIS retrievals. Each retrieved box covers 5×5 km^2 and is centered on the mean latitude and longitude of the aircraft cloud top segment nearest in time to the satellite over-

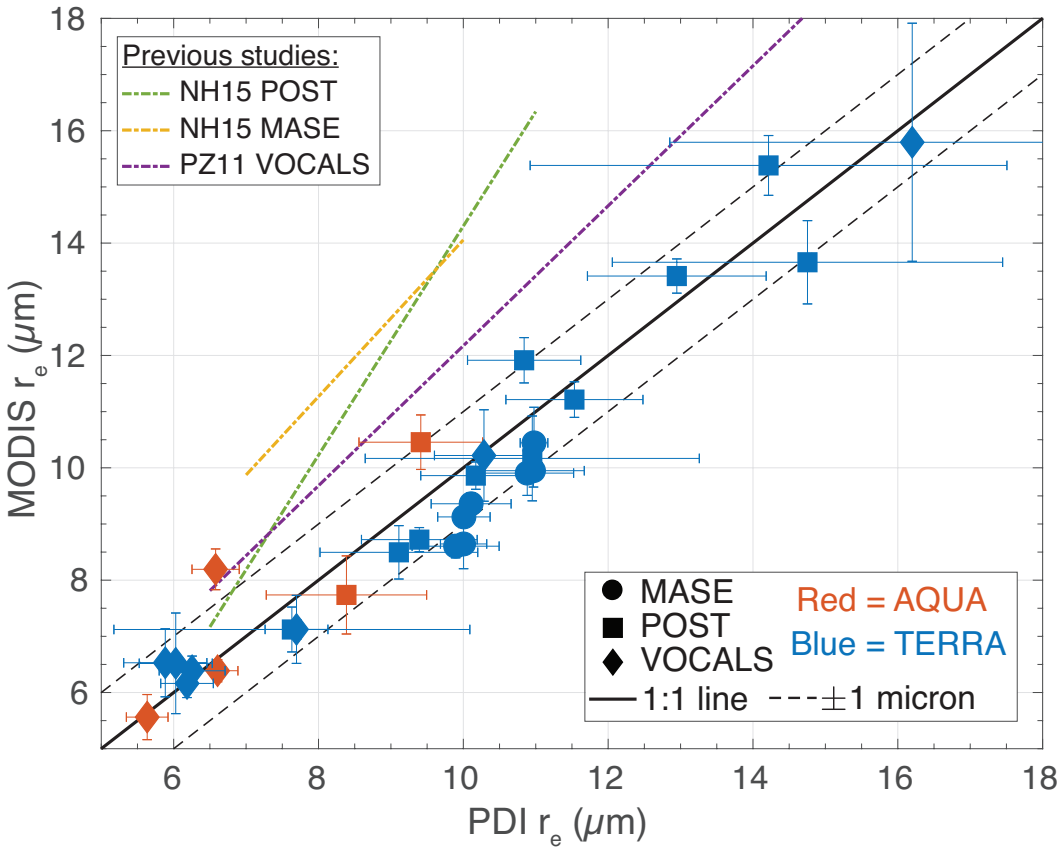

Figure 2. Comparison of MODIS r_e with flight mean cloud top PDI r_e. Vertical error bars are the standard deviation of MODIS r_e within the 5 × 5 km^2 box retrieved and horizontal error bars are $\sigma(r_{e,PDI})$ of the corresponding cloud top segment. Linear fits to results from previous studies are shown by dash-dot lines over the observed range of in situ values for each study. *Noble and Hudson* [2015] derive r_e from CAPS and *Painemal and Zuidema* [2011] use a Cloud Droplet Probe (CDP, see supporting information).

pass. No attempt is made to account for back trajectories, as *Min et al.* [2012] found negligible improvement in comparison statistics between projected and un-projected retrievals. The median time between aircraft sampling and satellite overpass is 30 minutes and the greatest time difference is 74 minutes. Box-mean r_e and H_σ are derived from pixels classified as fully overcast, single layer liquid cloud. Over the three field campaigns analyzed, a total of 29 MODIS retrievals correspond to Twin Otter cloud top flight segments. The use of near-cloud top legs and the top 30 m of sawtooth legs for the intercomparison is motivated by the fact that the 2.1 μm r_e retrieval is most sensitive to the level corresponding to $\tau \approx 1$, which typically occurs a few tens of meters below cloud base. Furthermore, the vertical gradient of r_e in this region is small, such that aircraft r_e is insensitive to leg-scale variability in cloud top height (see supporting information Figs. S1-S4). More detailed information on the intercomparison methodology can be found in the supporting information [*Miller et al.*, 2016; *Platnick*, 2000; *Zhang et al.*, 2017].

3 Results and discussion

In contrast with the findings of previous studies, MODIS and PDI r_e agree within uncertainty, i.e. there is no significant bias (Fig. 2). Mean bias ($\overline{\Delta r_e} = \overline{r_{e,MODIS} - r_{e,PDI}}$) is

Campaign	No. retrievals	$\overline{\Delta r_e}$ (μm)	$\overline{\lvert\Delta r_e\rvert}$ (μm)	Slope	R^2
All	29	-0.22	0.68	0.94	0.92
MASE	7	-0.97	0.97	1.28	0.87
POST	12	-0.10	0.73	1.08	0.96
VOCALS	10	0.15	0.42	0.92	0.96

Table 1. Intercomparison statistics for MASE, POST and VOCALS. Bias $\Delta r_e = r_{e,MODIS} - r_{e,PDI}$. Slope and R^2 are for linear least squares fits.

-0.22 μm, and mean absolute difference ($\overline{|\Delta r_e|}$) is 0.68 μm (Table 1), both well within uncertainty. It is particularly notable that the four points with the largest r_e fall within uncertainty of the one-to-one line, as past intercomparisons with aircraft observations have consistently found that bias increases with r_e [*Painemal and Zuidema*, 2011; *Noble and Hudson*, 2015; *Min et al.*, 2012; *King et al.*, 2013]. A more favorable comparison is to be expected for small r_e cases ($r_e \lesssim 8$ μm) since these cases best align with the DSD assumptions made by the MODIS r_e algorithm: a unimodal distribution, minimally variable in space. But these assumptions may not hold for the largest r_e cases, which are in drizzling conditions (for cases with $r_e > 12$ μm, $\overline{LWC_{CIP}} = 0.13$ g m^{-3}) and thus exhibit greater variability in distribution shape. This is consistent with *Zhang et al.* [2012], who find that simulated MODIS r_e retrievals from large eddy simulations of marine stratocumulus are negligibly affected by the presence of drizzle. Our results indicate MODIS retrievals are accurate (within 0.7 μm) and exhibit little bias for the range $5 < r_e < 16$ μm in stratocumulus.

To understand why we find agreement between MODIS and in situ measurements while previous studies did not, we examined a number of factors. We first describe those factors that we conclude are *not* significant contributors to MODIS retrieval bias in the relatively homogeneous, optically thick clouds sampled for this study. We find no correlation between Δr_e and spatial heterogeneity as measured by H_σ, likely because all retrieved scene-average H_σ values are less than 0.3, a threshold for significant inhomogeneity suggested by *Zhang and Platnick* [2011]. One idea that has been proposed [*Glienke et al.*, 2017] in a marine shallow convection setting is that the concentration of drops in the size range $40 < d < 80$ μm is not correctly measured by the most commonly deployed cloud probe combinations (i.e. CAPS or a similar combination such as CDP and 2DC/CIP). The PDI, however, is able to measure these drops, and we find that the data presented here do not exhibit a strong dependence on drops in this size range. This makes sense because MODIS retrievals of 2.1 μm r_e typically sample the top $\sim$ 100 m of cloud (see Supplementary Material) where drizzle drops are rare and do not contribute significantly to r_e. Drops in this size range could play a more important role for shorter wavelength r_e retrievals that penetrate deeper into cloud (i.e. 1.6 μm r_e) and in other cloud types like shallow cumulus that are deeper and more likely to form precipitation. More quantitatively, we find that PDI r_e correlates closely with the 90th percentile drop size from number concentration ($r^2 = 0.99$), which is typically between $13 < d < 26$ μm in this dataset and rarely exceeds $d = 40$ μm. This implies that larger drops simply do not occur in sufficient abundance near cloud top to dominate r_e in marine stratocumulus.

Two previous studies compared aircraft and MODIS r_e using Twin Otter instrumentation other than the PDI: *Noble and Hudson* [2015] analyzed CAPS data from POST and *Zheng et al.* [2011] analyzed PVM measurements from VOCALS. We independently reanalyze all CAPS and PVM in situ observations taken during MASE, POST and VOCALS to compare with PDI. In total, the merged database contains over 200,000 1 Hz observations in cloud (here defined as $r_e > 0.1$ μm) across the three campaigns. A comparison of all PDI r_e measurements with CAPS and PVM is shown in Fig. S3 in the Supplementary Material. Microphysical measurements from different probes on an aircraft should agree in the mean even though instantaneous values may not agree because different air volumes are sampled

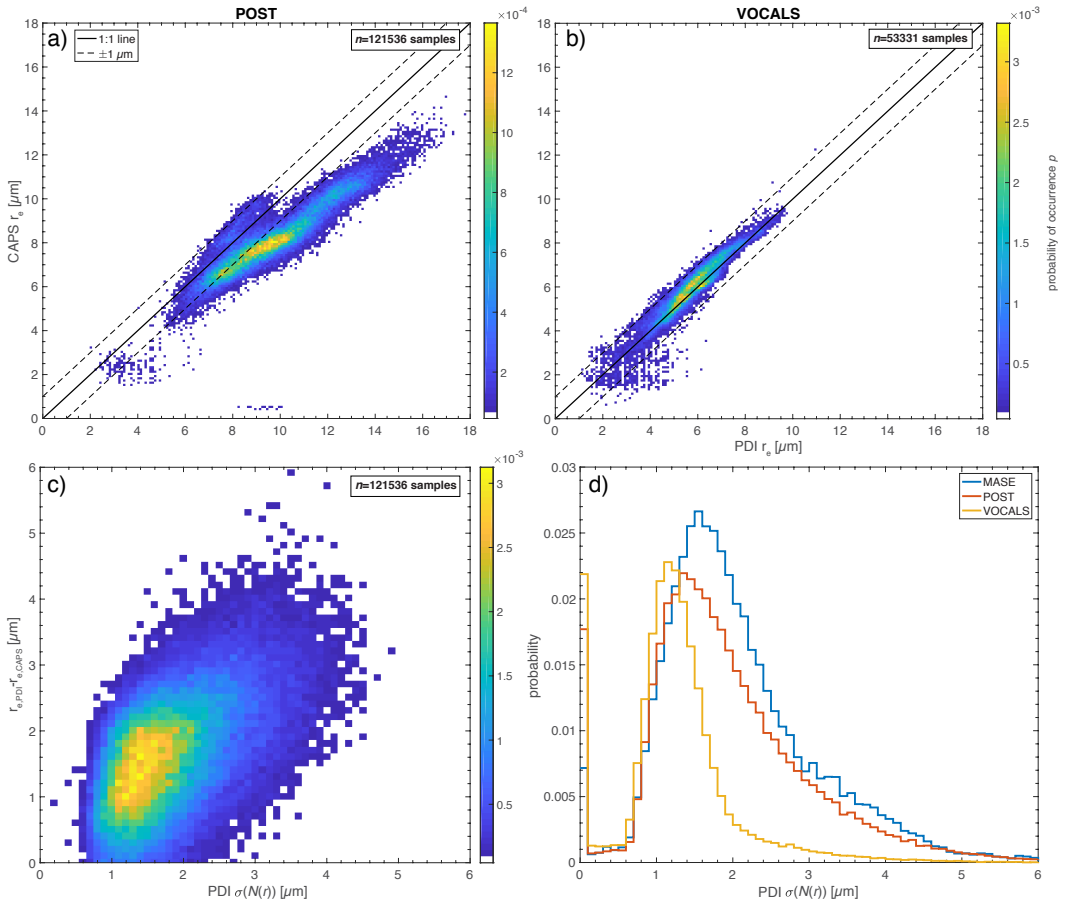

Figure 3. Top row: joint pdfs of PDI and CAPS r_e for POST (panel a) and VOCALS (panel b). Number of samples is the number of observations for which both probes measured $r_e > 0.1$ μm. Panel c: joint pdf of PDI DSD standard deviation σ and difference between PDI and CAPS r_e during POST. Number of samples is number of observations where $r_{e,PDI} - r_{e,CAPS} > 0.1$ μm. Panel d: pdfs of σ by field campaign. Joint pdf and histogram bin size is 0.1 μm for all variables in all panels and joint pdf values lower than 5×10^{-5} are not shown for visual clarity.

by each. Differences in mean values can thus be interpreted as being caused by a combination of differences in operating principles, calibration and/or drop size range measured.

During VOCALS, PDI and CAPS r_e are highly correlated and agree with each other quite well (a robust linear fit gives slope = 1.048 ± 0.001, $r^2 = 0.96$; Fig. 3, panel b). The TO primarily sampled polluted clouds during VOCALS that exhibited high number concentrations, small drop sizes and narrow DSDs. During POST, PDI and CAPS r_e are still highly correlated but there is a consistent offset of 1 to 2 μm toward larger PDI r_e (robust linear fit $r^2 = 0.88$; Fig. 3, panel a). Clouds during POST exhibit lower drop concentrations and broader DSDs. Recall that CAPS measurements are unavailable for MASE.

It is unlikely that calibration issues can explain the difference in relationships in derived r_e between POST and VOCALS. PDI calibration was verified using TO true air speed measurements and CAPS probes were routinely calibrated using glass beads during both experiments (see Fig. S4 for examples of CAS, CIP and PDI DSDs).

Instead, we find that the difference between PDI and CAPS r_e during POST varies with the breadth of the DSD, as measured by the standard deviation of the drop number distribution (Fig. 3, panel c). The distribution of σ during VOCALS is much narrower than MASE or POST, with values of $\sigma > 2$ μm seldom observed during VOCALS (Fig. 3, panel d). Given that PDI and the CAS component of CAPS have comparable sample volumes [*Baumgardner et al.*, 2011], we therefore speculate that CAPS may exhibit biases in the presence of broad DSDs and/or large drops, and that this is a primary cause of bias between CAPS and PDI-derived r_e.

4 Conclusions

In summary, we have performed an intercomparison of MODIS and PDI-derived r_e encompassing three field campaigns (MASE, POST and VOCALS). We show that PDI r_e agrees with MODIS r_e within uncertainty ($\overline{\Delta r_e} = -0.22$ μm, $\overline{|\Delta r_e|} = 0.68$ μm) for $5 < r_e < 16$ μm in marine stratocumulus, spanning non-drizzling to drizzling conditions. Good agreement between PDI and PVM LWC throughout the three field campaigns (Fig. 1) and between CAPS and PDI r_e in a subset of conditions during VOCALS (Fig. 3) provide mutual reinforcement for the PDI-MODIS intercomparison.

Our results contradict those of previous studies [*Painemal and Zuidema*, 2011; *Zheng et al.*, 2011; *Noble and Hudson*, 2015; *Min et al.*, 2012; *King et al.*, 2013] that have found that in situ r_e is smaller than that retrieved by satellite, leading those authors to conclude that MODIS systematically overestimates r_e. Improvements in MODIS retrieval algorithms from Collections 5.1 to 6 can account for up to $\sim$ 1 μm of the previously observed bias [*Rausch et al.*, 2017], but have little power to explain the observed biases of $2 - 5$ μm found by past intercomparisons. We suggest that estimating the contribution of instrumentation error to r_e bias is not simply a matter of calculating sampling error; there is also intrinsic error associated with the drop size range effectively measured by the instrumentation employed. By comparing PDI measurements with those from the probes used in previous studies (CAPS and PVM) during MASE, POST and VOCALS, we have demonstrated that CAPS and PDI agree well in small r_e conditions, but PDI typically measures larger r_e than CAPS and PVM in the presence of broad DSDs, which are also associated with larger r_e. We note that the excellent agreement between PDI and MODIS and the plausible explanation proposed here to explain previous disagreement in the literature do not necessarily mean they are right.

Looking forward, the tight correlation of PDI r_e and 90th percentile drop size ($r^2 = 0.99$) may merit further consideration in parameterization development. It is curious that more extreme values/higher percentiles have lower correlations with r_e, and we interpret this as symptomatic of our finding that the largest drops do not occur in sufficiently large numbers to strongly influence total r_e in the near cloud top region relevant to MODIS retrievals. This study also highlights the need for a more comprehensive assessment of in situ cloud

probes, especially to compare PDI with other updated or novel instrumentation such as CDP-2 or HOLODEC (a holography-based probe) and to better understand the conditions in which the width of the DSD is consistently measured, as this appears to be the primary difference among the probes analyzed. Finally, further studies of MODIS r_e retrievals with appropriate instrumentation in other cloud types are needed to assess whether the results of this study are universally applicable.

Acknowledgments

We thank the Twin Otter pilots, crew and fellow scientists during MASE, POST and VOCALS as well as Aaron Bansemer for his insightful comments on an earlier version of the manuscript. This work was funded by the National Science Foundation under grant AGS-1139746. POST and VOCALS data can be freely obtained at data.eol.ucar.edu and MASE data for the flights used for this manuscript can be obtained at https://doi.org/10.5281/zenodo.1035928. MODIS Level 2 Collection 6 retrievals are available from https://ladsweb.modaps.eosdis.nasa.gov.

References

Albrecht, H.-E., N. Damaschke, M. Borys, and C. Tropea (2013), *Laser Doppler and Phase Doppler Measurement Techniques*, Experimental Fluid Mechanics, Springer Berlin Heidelberg.

Andersen, H., J. Cermak, J. Fuchs, R. Knutti, and U. Lohmann (2017), Understanding the drivers of marine liquid-water occurrence and properties with global observations using neural networks, *Atmos. Chem. Phys.*, *17*, 9535–9546, doi:10.594/acp-17-9535-2017.

Bachalo, W. (2000), Spray diagnostics for the twenty-first century, *Atomization and Sprays*, *10*(3-5), 439–474, doi:10.1615/AtomizSpr.v10.i3-5.110.

Baum, B. A., W. P. Menzel, R. A. Frey, D. C. Tobin, R. E. Holz, S. A. Ackerman, A. K. Heidinger, and P. Yang (2012), MODIS cloud-top property refinements for collection 6, *J. Appl. Meteor. Climatol.*, *51*, 1145–1163, doi:10.1176/JAMC-D-11-0203.1.

Baumgardner, D., H. Jonsson, W. Dawson, D. O'Connor, and R. Newton (2001), The cloud, aerosol and precipitation spectrometer: a new instrument for cloud investigations, *Atmos. Res.*, *59–60*, 251–264, doi:10.1016/S0169-8095(01)00119-3.

Baumgardner, D., J. L. Brenguier, A. Bucholtz, H. Coe, P. DeMott, T. J. Garrett, J. F. Gayet, M. Hermann, A. Heymsfield, A. Korolev, M. Krämer, A. Petzold, W. Strapp, P. Pilewskie, J. Taylor, C. Twohy, M. Wendisch, W. Bachalo, and P. Chuang (2011), Airborne instruments to measure atmospheric aerosol particles, clouds and radiation: A cook's tour of mature and emerging technology, *Atmos. Res.*, *102*, 10–29, doi:10.1016/j.atmosres.2011.06.021.

Bennartz, R., and J. Rausch (2017), Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations, *Atmos. Chem. Phys.*, *17*, 9815–9836, doi:10.5194/acp-17-9815-2017.

Carman, J. K., D. L. Rossiter, D. Khelif, H. H. Jonsson, I. C. Faloona, and P. Y. Chuang (2012), Observational constraints on entrainment and the entrainment interface layer in stratocumulus, *Atmos. Chem. Phys.*, *12*, 11,135–11,152, doi:10.5194/acp-12-11135-2012.

Chandrakar, K. K., W. Cantrell, K. Chang, D. Ciochetto, D. Niedermeier, M. Ovchinnikov, R. A. Shaw, and F. Yang (2016), Aerosol indirect effect from turbulence-induced broadening of cloud-droplet size distributions, *Proc. Natl. Acad. Sci. USA*, *113*(50), 14,243–14,248, doi:10.1073/pnas.1612686113.

Chang, K., J. Bench, M. Brege, W. Cantrell, K. Chandrakar, D. Ciochetto, C. Mazzoleni, L. R. Mazzoleni, D. Niedermeier, and R. A. Shaw (2016), A laboratory facility to study gas-aerosol-cloud interactions in a turbulent environment: The Π chamber, *Bull. Amer. Meteor. Soc.*, *97*, 2343–2358, doi:10.1175/BAMS-D-15-00203.1.

Chuang, P. Y., E. W. Saw, J. D. Small, R. A. Shaw, C. M. Sipperley, G. A. Payne, and W. D. Bachalo (2008), Airborne phase Doppler interferometry for cloud microphysical measurements, *Aerosol Sci. Technol.*, *42*, 685–703, doi:10.1080/02786820802232956.

Ditas, F., R. A. Shaw, H. Siebert, M. Simmel, B. Wehner, and A. Wiedensohler (2012), Aerosols-cloud microphysics-thermodynamics-turbulence: evaluating supersaturation in a marine stratocumulus cloud, *Atmos. Chem. Phys.*, *12*, 2459–2468, doi:10.5194/acp-12-2459-2012.

Gerber, H. (1991), Direct measurement of suspended particulate volume concentration and far-infrared extinction coefficient with a laser diffraction instrument, *Appl. Opt.*, *30*, 4824–4831, doi:10.1364/AO.30.004824.

Gerber, H., B. G. Arends, and A. S. Ackerman (1994), New microphysics sensor for aircraft use, *Atmos. Res.*, *31*, 235–252, doi:10.1016/0169-8095(94)90001-9.

Gerber, H., G. Frick, S. P. Malinowski, H. Jonsson, D. Khelif, and S. K. Krueger (2013), Entrainment rates and microphysics in POST stratocumulus, *J. Geophys. Res. Atmos.*, *118*, 12,094–12,109, doi:10.1002/jgrd.50878.

Glienke, S., A. Kostinski, J. Fugal, R. A. Shaw, S. Borrmann, and J. Stith (2017), Cloud droplets to drizzle: Contribution of transition drops to microphysical and optical properties of marine stratocumulus clouds, *Geophys. Res. Lett.*, *44*, doi:10.1002/2017GL074430.

Grosvenor, D. P., and R. Wood (2014), The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals within marine liquid water clouds, *Atmos. Chem. Phys.*, *14*, 7291–7321, doi:10.5194/acp-14-7291-2014.

Henrich, F., H. Siebert, E. Jäkel, R. A. Shaw, and M. Wendisch (2010), Collocate measurements of boundary layer cloud microphysical and radiative properties: A feasibility study, *J. Geophys. Res.*, *115*, D24,214, doi:10.1029/2010JD013930.

Kato, S., and A. Marshak (2009), Solar zenith and viewing geometry-dependent errors in satellite retrieved cloud optical thickness: Marine stratocumulus case, *J. Geophys. Res.*, *114*, D01,202, doi:10.1029/2008JD010579.

King, N. J., K. N. Bower, J. Crosier, and I. Crawford (2013), Evaluating MODIS cloud retrievals with in situ observations from VOCALS-REx, *Atmos. Chem. Phys.*, *13*, 191–209, doi:10.5194/acp-13-191-2013.

Kokhanovsky, A. A., D. Painemal, and V. V. Rozanov (2013), The intercomparison of satellite-derived and in situ profiles of droplet effective radii in marine stratocumulus clouds, *IEEE Geoscience and Remote Sensing Letters*, *10*(5), 1147–1151, doi:10.1109/LGRS.2012.2233710.

Lance, S., C. A. Brock, D. Rogers, and J. A. Gordon (2010), Water droplet calibration of the Cloud Droplet Probe (CDP) and in-flight performance in liquid, ice and mixed-phase clouds during ARCPAC, *Atmos. Meas. Tech.*, *3*, 1683–1706, doi:10.5194/amt-3-1683-2010.

Liang, L., L. D. Girolamo, and W. Sun (2015), Bias in MODIS cloud drop effective radius for oceanic water clouds as deduced from optical thickness variability across scattering angles, *J. Geophys. Res. Atmos.*, *120*, 7661–7681, doi:10.1002/2015JD023256.

Lu, M.-L., W. C. Conant, H. H. Jonsson, V. Varutbangkul, R. C. Flagan, and J. H. Seinfeld (2007), The Marine Stratus/Stratocumulus Experiment (MASE): Aerosol-cloud relationships in marine stratocumulus, *J. Geophys. Res.*, *112*, D10,209, doi:10.1029/2006JD007985.

Maddux, B. C., S. A. Ackerman, and S. Platnick (2010), Viewing geometry dependencies in MODIS cloud products, *J. Atmos. Oceanic Technol.*, *27*, 1519–1528, doi:10.1175/2010JTECHA1432.1.

Marshak, A. S., S. Platnick, T. Várnai, G. Wen, and R. F. Cahalan (2006), Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes, *J. Geophys. Res.*, *111*, D09,207, doi:10.1029/2005JD006686.

Martin, G. M., D. W. Johnson, and A. Spice (1994), The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds, *J. Atmos. Sci.*, *51*(13), 1823–1842, doi:10.1175/1520-0469(1994)051<1823:TMAPOE>2.0.CO;2.

Mechoso, C. R., R. Wood, R. Weller, C. S. Bretherton, A. D. Clarke, H. Coe, C. Fairall, J. T. Farrar, G. Feingold, R. Garreaud, C. Grados, J. McWilliams, S. P. de Szoeke, S. E. Yuter, and P. Zuidema (2014), Ocean-Cloud-Atmosphere-Land interactions in the South-

eastern Pacific: The VOCALS program, *Bull. Amer. Meteor. Soc.*, *95*, 357–375, doi:10.1175/BAMS-D-11-00246.1.

Miller, D. J., Z. Zhang, A. S. Ackerman, S. Platnick, and B. A. Baum (2016), The impact of cloud vertical profile on liquid water path retrieval based on the bispectral method: A theoretical study based on large-eddy simulations of shallow marine boundary layer clouds, *J. Geophys. Res. Atmos.*, *121*, 4122–4141, doi:10.1002/2015JD024322.

Min, Q., E. Joseph, Y. Lin, L. Min, B. Yin, P. H. Daum, L. I. Kleinman, J. Wang, and Y.-N. Lee (2012), Comparison of MODIS cloud microphysical properties with in-situ measurements over the Southeast Pacfic, *Atmos. Chem. Phys.*, pp. 11,261–11,273, doi:10.5194/acp-12-11261-2012.

Nakajima, T., M. D. King, J. D. Spinhirne, and L. F. Radke (1991), Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. part II: marine stratocumulus observations, *J. Atmos. Sci.*, *48*(5), 728–750, doi:10.1175/1520-0469(1991)048<0728:DOTOTA>2.0.CO;2.

Nakajima, T. Y., K. Suzuki, and G. L. Stephens (2010), Droplet growth in warm water clouds observed by the A-Train. part I: Sensitivity analysis of the MODIS-derived cloud droplet sizes, *J. Atmos. Sci.*, *67*, 1884–1896, doi:10.1175/2009JAS3280.1.

Noble, S. R., and J. G. Hudson (2015), MODIS comparisons with northeastern Pacific in situ stratocumulus microphysics, *J. Geophys. Res. Atmos.*, *120*, 8332–8344, doi:10.1102/2014JD0022785.

Painemal, D., and P. Zuidema (2011), Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements, *J. Geophys. Res.*, *116*, D24,206, doi:10.1029/2011JD016155.

Pawlowska, H., W. W. Grabowski, and J.-L. Brenguier (2006), Observations of the width of cloud droplet spectra in stratocumulus, *Geophys. Res. Lett.*, *33*, L19,810, doi:10.1029/2006GL026841.

Pinnick, R. G., D. M. Garvey, and L. D. Duncan (1981), Calibration of Knollenberg FSSP light-scattering counters for measurement of cloud droplets, *J. Appl. Meteorol.*, *20*, 1049–1057, doi:10.1175/1520-0450(1981)020<1049:COKFLS>2.0.CO;2.

Platnick, S., and F. P. J. Valero (1995), A validation of a satellite cloud retrieval during ASTEX, *J. Atmos. Sci.*, *52*(16), 2985–3001, doi:10.1175/1520-0469(1995)052<2985:AVOASC>2.0.CO;2.

S. Platnick (2000), Vertical photon transport in cloud remote sensing problems, *J. Geophys. Res.*, *105*, 22919–22935, doi:10.1029/2000JD900333.

Platnick, S., K. G. Meyer, M. D. King, G. Wind, N. Amarasinghe, B. Marchant, G. T. Arnold, Z. Zhang, P. A. Hubanks, R. E. Holz, P. Yang, W. L. Ridgway, and J. Riedi (2017), The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua, *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525, doi:10.1109/TGRS.2016.2610522.

Quaas, J., Y. Ming, S. Menon, T. Takemura, M. Wang, J. E. Penner, A. Gettelman, U. Lohmann, N. Bellouin, O. Boucher, A. M. Sayer, G. E. Thomas, A. McComiskey, G. Feingold, C. Hoose, J. E. Kristjánsson, X. Liu, Y. Balkanski, L. J. Donner, P. A. Ginoux, P. Stier, B. Grandey, J. Feichter, I. Sednev, S. E. Bauer, D. Koch, R. G. Grainger, A. Kirkevåg, T. Iversen, Ø. Seland, R. Easter, S. J. Ghan, P. J. Rasch, H. Morrison, J.-F. Lamarque, M. J. Iacono, S. Kinne, and M. Schulz (2009), Aerosol indirect effects - general circulation model intercomparison and evaluation with satellite data, *Atmos. Chem. Phys.*, *9*, 8697–8717, doi:10.5194/acp-9-8697-2009.

Rausch, J., K. Meyer, R. Bennartz, and S. Platnick (2017), Differences in liquid cloud droplet effective radius and number concentration estimates between MODIS collections 5.1 and 6 over global oceans, *Atmos. Meas. Tech.*, *10*, 2105–2116, doi:10.5194/amt-10-2105-2017.

Ruehl, C. R., P. Y. Chuang, and A. Nenes (2008), How quickly do cloud droplets form on atmospheric particles?, *Atmos. Chem. Phys.*, *8*, 1043–1055, doi:10.5194/acp-8-1043-2008.

Ruehl, C. R., P. Y. Chuang, and A. Nenes (2010), Aerosol hygroscopicity at high (99 to 100%) relative humidities, *Atmos. Chem. Phys.*, *10*, 1329–1344, doi:10.5194/acp-10-1329-

2010.

Ruehl, C. R., P. Y. Chuang, A. Nenes, C. D. Kappa, K. R. Kolesar, and A. H. Goldstein (2012), Strong evidence of surface tension reduction in microscopic aqueous droplets, *Geophys. Res. Lett.*, *39*, L23,801, doi:10.1029/2012GL053706.

Saw, E.-W., R. A. Shaw, J. P. L. C. Salazar, and L. R. Collins (2012), Spatial clustering of polydisperse inertial particles in turbulence: II. comparing simulation with experiment, *New J. Phys.*, *14*(10), 105,013, doi:10.1088/1367-2630/14/10/105031.

Siebert, H., and R. A. Shaw (2017), Supersaturation fluctuations during the early stage of cumulus formation, *J. Atmos. Sci.*, *74*(4), 975–988, doi:10.1175/JAS-D-16-0115.1.

Siebert, H., H. Franke, K. Lehmann, R. Maser, E. W. Saw, D. Schell, R. A. Shaw, and M. Wendisch (2006), Probing finescale dynamics and microphysics of clouds with helicopter-borne measurements, *Bull. Amer. Meteor. Soc.*, *87*, 1727–1738, doi:10.1175/BAMS-87-I2-I727.

Siebert, H., M. Beals, J. Bethke, E. Bierwirth, T. Conrath, K. Dieckmann, F. Ditas, A. Ehrlich, D. Farrell, S. Hartmann, M. A. Izaguirre, J. Katzwinkel, L. Nuijens, G. Roberts, M. Schäfer, R. A. Shaw, T. Schmeissner, I. Serikov, B. Stevens, F. Stratmann, B. Wehner, M. Wendisch, F. Werner, and H. Wex (2013), The fine-scale structure of the trade wind cumuli over Barbados – an introduction to the CARRIBA project, *Atmos. Chem. Phys.*, *13*, 10,061–10,077, doi:10.5194/acp-13-10061-2013.

Siebert, H., R. A. Shaw, J. Ditas, T. Schmeissner, S. P. Malinowski, E. Bodenschatz, and H. Xu (2015), High-resolution measurement of cloud microphysics and turbulence at a mountaintop station, *Atmos. Meas. Tech.*, *8*, 3219–3228, doi:10.5194/amt-8-3219-2015.

Su, W., N. G. Loeb, K.-M. Xu, G. L. Schuster, and Z. A. Eitzen (2010), An estimate of aerosol indirect effect from satellite measurements with concurrent meteorological analysis, *J. Geophys. Res.*, *115*, D18,219, doi:10.1029/2010JD013948.

Tropea, C. (2011), Optical particle characterization in flows, *Annu. Rev. Fluid Mech.*, *43*(1), 399–426.

Wang, J., P. H. Daum, S. S. Yum, Y. Liu, G. I. Senum, M.-L. Lu, J. H. Seinfeld, and H. Jonsson (2009), Observations of marine stratocumulus microphysics and implications for processes controlling droplet spectra: Results from the Marine Stratus/Stratocumulus Experiment, *J. Geophys. Res.*, *114*, D18,210, doi:10.1029/2008JD011035.

Wendisch, M., T. J. Garrett, and J. W. Strapp (2002), Wind tunnel tests of the airborne PVM-100A response to large droplets, *J. Atmos. Oceanic Technol.*, *19*, 1577–1584, doi:10.1175/1520-0426(2002)019<1577:WTTOTA>2.0.CO;2.

Yuan, T., L. A. Remer, and Y. Hu (2011), Microphysical, macrophysical and radiative signatures of volcanic aerosols in trade wind cumulus observed by the A-Train, *Atmos. Chem. Phys.*, *11*, 7119–7132, doi:10.5194/acp-11-7119-2011.

Zhang, Z. and S. Platnick (2011), An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands, *J. Geophys. Res*, *116*, D20,215, doi:10.1029/2011JD016216.

Zhang, Z., A. S. Ackerman, G. Feingold, S. Platnick, R. Pincus, and H. Xue (2012), Effects of cloud horizontal inhomogeneity and drizzle on remote sensing of cloud droplet effective radius: Case studies based on large-eddy simulations, *J. Geophys. Res.*, *117*, D19,208, doi:10.1029/2012JD07655.

Zhang, Z., X. Dong, B. Xi, H. Song, P.-L. Ma, S. J. Ghan, S. Platnick, and P. Minnis (2017), Intercomparisons of marine boundary layer cloud properties from the ARM CAP-MBL campaign and two MODIS cloud products, *J. Geophys. Res. Atmos.*, *122*, 2351–2365, doi:10.1002/2016JD025763.

Zheng, X., B. Albrecht, H. H. Jonsson, D. Khelif, G. Feingold, P. Minnis, J. K. Ayers, P. Chuang, S. Donaher, D. Rossiter, V. Ghate, J. Ruiz-Plancarte, and S. Sun-Mack (2011), Observations of the boundary layer, cloud, and aerosol variability in the southeast Pacific near-coastal marine stratocumulus during VOCALS-REx, *Atmos. Chem. Phys.*, *11*, 9943–9959, doi:10.5194/acp-11-9943-2011.